Statistically Efficient Methods for Pitch and DOA Estimation

Jensen, Jesper Rindom; Christensen, Mads Græsbøll; Jensen, Søren Holdt

Published in:
2013 IEEE International Conference on Acoustics, Speech, and Signal Processing

DOI (link to publication from Publisher):
10.1109/ICASSP.2013.6638389

Publication date:
2013

Document Version
Accepted author manuscript, peer reviewed version

STATISTICALLY EFFICIENT METHODS FOR PITCH AND DOA ESTIMATION

Jesper Rindom Jensen[†]*, Mads Græsbøll Christensen*[†]*, and Søren Holdt Jensen*[‡]

[†]Audio Analysis Lab, AD:MT
Aalborg University, Denmark
{jrj,mgc}@create.aau.dk

[‡]Dept. of Electronic Systems
Aalborg University, Denmark
shj@es.aau.dk

ABSTRACT

Traditionally, direction-of-arrival (DOA) and pitch estimation of multichannel, periodic sources have been considered as two separate problems. Separate estimation may render the task of resolving sources with similar DOA or pitch impossible, and it may decrease the estimation accuracy. Therefore, it was recently considered to estimate the DOA and pitch jointly. In this paper, we propose two novel methods for DOA and pitch estimation. They both yield maximum-likelihood estimates in white Gaussian noise scenarios, where the SNR may be different across channels, as opposed to state-of-the-art methods. The first method is a joint estimator, whereas the latter use a cascaded approach, but with a much lower computational complexity. The simulation results confirm that the proposed methods outperform state-of-the-art methods in terms of estimation accuracy in both synthetic and real-life signal scenarios.

Index Terms— Maximum likelihood, direction-of-arrival, pitch, joint estimation, closed-form estimates.

1. INTRODUCTION

In many applications involving quasi-periodic signals such as voiced speech or musical instrument recordings, an array of microphones are used for picking up the desired signal. Hearing-aids, hand-held devices, teleconference systems, and surveillance systems are just a few examples of such applications. Due to periodicity property and the multichannel recording scheme, the desired signal has a direction-of-arrival (DOA) and a pitch. These two parameters are some of the main paraphernalia in many of signal processing methods utilized in the abovementioned applications. For example, the DOA, the pitch, or even both parameters are necessary in various methods for tracking [1], source separation [2], enhancement [3–5], and compression [6, 7].

Due to the importance of knowing these parameters, several methods have been proposed for both DOA and pitch estimation. For multichannel speech and audio, DOA estimation has often been treated as a broadband problem and an overview of such methods can be found in, e.g., [8–12]. The pitch estimation problem has mainly been considered a single-channel estimation problem [13], however, in the recent years a few methods have been proposed for multichannel scenarios [14–16]. That is, the estimation of the DOA and pitch has traditionally been treated as two separate problems as hinted by the above references. Estimating the DOA and pitch separately may, however, cause two problems: the estimation accuracy will most likely be suboptimal, and it might not be possible to resolve sources overlapping in one dimension. Motivated by these observations and due to an increasing computational capability, it has therefore been considered recently to estimate the DOA and pitch jointly by assuming a spatio-temporal harmonic model for the desired, multichannel, periodic signal. In other words, the desired signal is assumed to consist of a number of narrowband signals with harmonically related carrier frequencies. Some examples of joint DOA and pitch estimation techniques are a maximum-likelihood based method [17], subspace methods [18–20], correlation-based methods [21, 22], filtering methods [23, 24], and nonlinear least squares (NLS) methods [25]. Note that some of the methods mentioned above consider time delay estimation and not DOA estimation, however, these parameters are closely related as clarified later.

In this paper, we propose two new methods for pitch and DOA estimation. Both of the proposed methods yield maximum likelihood (ML) estimates when the noise is white Gaussian in each channel even when the signal-to-noise ratios and the attenuations of the periodic signal are different across the channels. This is not the case for any of the joint pitch and DOA estimators in [17–25]. The first of the proposed methods directly and jointly maximizes the likelihood function for the pitch and the DOA. This involves a two-dimensional search, i.e., the computational complexity will be high. We therefore also propose a cascaded method, where we first obtain a ML estimate of the pitch using the method in [16]. Then, in the second stage, a closed-form estimate of the DOA involving weighted least-squares (WLS) is obtained.

The remainder of the paper is organized as follows: first, the signal model and the problem formulation considered in the paper are defined in Section 2. Then, we propose a joint and a cascaded pitch and DOA estimator in Section 3 and Section 4, respectively. In Section 5, we present the experimental results and, finally, we relate our work to the state of the art in Section 6.

2. PROBLEM FORMULATION

To facilitate the derivation of statistically efficient pitch and DOA estimators, we first present the multichannel model under consideration. In the proposed methods, it is assumed that N data snapshots have been obtained using K sensors. The data obtained using sensor k can be represented by a data vector $\mathbf{x}_k(n) \in \mathbb{C}^N$ defined as

$$\mathbf{x}_k(n) = \begin{bmatrix} x_k(n) & x_k(n+1) & \cdots & x_k(n+N-1) \end{bmatrix}^T, \quad (1)$$

for $k = 0, \ldots, K-1$, where $x_k(n)$ is the signal observed at sensor k at time instance n. In scenarios with a single periodic source in an anechoic environment, we can model each of these vectors as

$$\mathbf{x}_k(n) = \beta_k \mathbf{Z}\mathbf{D}(k)\boldsymbol{\alpha} + \mathbf{e}_k(n), \quad (2)$$

with $\mathbf{Z} = \begin{bmatrix} \mathbf{z}(\omega_0) & \cdots & \mathbf{z}(L\omega_0) \end{bmatrix}$, $[\mathbf{z}(l\omega_0)]_n = e^{jln\omega_0}$ for $n = 0, \ldots, N-1$, $[\mathbf{D}(k)]_{ll} = e^{-jl\omega_0 f_s \tau_k}$ for $l = 1, \ldots, L$,

This work was supported in part by the Villum Foundation.

while all other entries of $\mathbf{D}(k)$ equals zero. Moreover, $\boldsymbol{\alpha} = \begin{bmatrix}\alpha_1 & \cdots & \alpha_L\end{bmatrix}^T$, β_k is the attenuation of the source wave from the source position to the position of sensor k, ω_0 is the pitch, L is the harmonic model order, f_s is the sampling frequency, τ_k is the delay of the source signal between sensor 0 and sensor k, and α_l is the complex amplitude of the lth harmonic. The factors β_k can be used for modeling, e.g, the attenuation of sound waves over distance as well as different sensor characteristics. Furthermore, the lth complex amplitude is given by $\alpha_l = A_l e^{j\phi_l}$ where $A_l > 0$ and ϕ_l is the real amplitude and the phase, respectively. When the array structure is known, the model can be detailed further by modeling the delay τ_k. In the remainder of the paper, we will assume a uniform linear array (ULA) structure. This enables us to model the delay from sensor k to sensor 0 as

$$\tau_k = kdc^{-1}\sin\theta, \tag{3}$$

where d is the spacing between two consecutive sensors in the ULA, $\theta \in [-90^\circ; 90^\circ]$ is the (DOA) of the source wave onto the ULA, and c is the wave propagation speed. Finally, we assume that the noise is white Gaussian in each channel, and that the noise is uncorrelated across channels. We also assume that the variance of the noise in each channel σ_k^2 is different. While these noise assumptions may appear limiting in practice, they are the best choices for the noise probability density function (pdf) when nothing about it is known, since the white Gaussian noise distribution can be shown to maximize the entropy of the noise [26].

Under these assumptions, the likelihood function for the complex data vector $\mathbf{x}_k(n)$ can be written as

$$p(\mathbf{x}_k(n); \boldsymbol{\psi}) = (\pi\sigma_k^2)^{-N} e^{-\frac{1}{\sigma_k^2}\|\mathbf{e}_k(n)\|^2}, \tag{4}$$

with $\boldsymbol{\psi}$ being the vector containing the signal parameters. Assuming the deterministic part of the model in (2) is stationary, we can write the likelihood for the whole set of data vectors $\{\mathbf{x}_k(n)\}_{k=0}^{K-1}$ as

$$\begin{aligned} p(\{\mathbf{x}_k(n)\}; \{\boldsymbol{\psi}\}) &= \prod_{k=0}^{K-1} p(\mathbf{x}_k(n); \boldsymbol{\psi}) \qquad (5)\\ &= \frac{1}{\pi^{NK}\left(\prod_{k=0}^{K-1}\sigma_k^2\right)^N} e^{-\sum_{k=0}^{K-1}\frac{1}{\sigma_k^2}\|\mathbf{e}_k(n)\|^2}. \end{aligned}$$

By taking the logarithm of (5) we get the log-likelihood function

$$\ln p(\{\mathbf{x}_k(n)\}; \{\boldsymbol{\psi}\}) = -NK\ln\pi - N\sum_{k=0}^{K-1}\ln\sigma_k^2 - \sum_{k=0}^{K-1}\frac{\|\mathbf{e}_k(n)\|^2}{\sigma_k^2}. \tag{6}$$

The task is then to obtain statistically efficient estimates of the pitch ω_0 and the DOA θ given N data snapshots from K sensors by maximizing the likelihood.

3. JOINT PITCH AND DOA ESTIMATION

Utilizing the aforementioned likelihood functions, we derive the proposed joint pitch and DOA estimation method in this section. To achieve this, we maximize the log-likelihood in (6) wrt. these parameters. First, we differentiate the log-likelihood wrt. to the unknown complex amplitudes $\boldsymbol{\alpha}$ and equate with zero to obtain the following amplitude estimates

$$\hat{\boldsymbol{\alpha}} = \left[\sum_{k=0}^{K-1}\frac{\beta_k^2}{\sigma_k^2}\mathbf{D}^H(k)\mathbf{Z}^H\mathbf{Z}\mathbf{D}(k)\right]^{-1}\sum_{k=0}^{K-1}\frac{\beta_k}{\sigma_k^2}\mathbf{D}^H(k)\mathbf{Z}^H\mathbf{x}_k(n). \tag{7}$$

Using a similar procedure, we can find an estimate of the attenuation β_k of the periodic source on the kth sensor as

$$\hat{\beta}_k = \frac{\mathrm{Re}\left\{\boldsymbol{\alpha}^H\mathbf{D}^H(k)\mathbf{Z}^H\mathbf{x}_k(n)\right\}}{\boldsymbol{\alpha}^H\mathbf{D}^H(k)\mathbf{Z}^H\mathbf{Z}\mathbf{D}(k)\boldsymbol{\alpha}} \tag{8}$$

where $\mathrm{Re}\{\cdot\}$ denotes the real part of a complex number. Finally, we differentiate the log-likelihood wrt. the noise variance on sensor k, equate with zero, and solve for the variance which yields

$$\hat{\sigma}_k^2 = N^{-1}\|\hat{\mathbf{e}}_k(n)\|^2, \tag{9}$$

with $\hat{\mathbf{e}}_k(n) = \mathbf{x}(n) - \hat{\beta}_k\mathbf{Z}\mathbf{D}(k)\hat{\boldsymbol{\alpha}}$. Apparently, the amplitude, attenuation factor and noise variance estimates are dependent on each other. We therefore propose to estimate these iteratively using the expressions in (7), (8), and (9), where the β_k's and σ_k^2's are initially set to 1. The log-likelihood is convex wrt. the product $\beta_k\boldsymbol{\alpha}$, and the algorithm will therefore converge [27]. According to our experience, three iterations typically sufficient. Note that the estimates of $\boldsymbol{\alpha}$ and β_k are not unique, but the product of the two is.

Then, we can combine (6) and (9) to obtain the concentrated log-likelihood for our set of data vectors, which depends only on the pitch ω_0 and the DOA θ. This yields

$$\ln p(\{\mathbf{x}_k(n)\}; \omega_0, \theta) = -NK(1+\ln\pi) - N\sum_{k=0}^{K-1}\ln\hat{\sigma}_k^2. \tag{10}$$

The joint maximum likelihood estimator (MLE) of the pitch and the DOA is then given by

$$\left\{\hat{\omega}_0, \hat{\theta}\right\} = \arg\min_{\{\omega_0,\theta\}\in\Omega_0\times\Theta}\sum_{k=0}^{K-1}\ln\left\|\mathbf{x}_k(n) - \hat{\beta}_k\mathbf{Z}\mathbf{D}(k)\hat{\boldsymbol{\alpha}}\right\|^2, \tag{11}$$

where Ω_0 and Θ are sets of candidate fundamental frequencies and DOAs, respectively. We denominate this estimator the joint maximum likelihood (JML) method. Estimating the pitch and DOA jointly is beneficial in scenarios with multiple periodic sources present, as we might be able to resolve sources with similar pitch as long as they are sufficiently spaced in DOA and vice versa.

4. CLOSED-FORM DOA ESTIMATION

While joint pitch and DOA estimation is advantageous in terms of resolving overlapping sources, it has the disadvantage of a high computational complexity since a two-dimensional search is required. To alleviate this, we also propose a cascaded pitch and DOA estimation method. First, the pitch is estimated using the multichannel pitch estimator proposed in [16]. Then, as we show in this section, we can obtain a statistically efficient estimate of the DOA by utilizing the ML pitch estimate.

To obtain such a DOA estimate, we first obtain a least squares (LS) estimate of the attenuated complex amplitudes $\boldsymbol{\alpha}'(k) = \beta_k\boldsymbol{\alpha}(k)$ on each channel as

$$\hat{\boldsymbol{\alpha}}'(k) = \arg\min_{\boldsymbol{\alpha}'}\|\mathbf{x}_k(n) - \mathbf{Z}\boldsymbol{\alpha}'(k)\|^2, \tag{12}$$

where $\boldsymbol{\alpha'}(k) = \beta_k \begin{bmatrix}\alpha_1(k) & \cdots & \alpha_L(k)\end{bmatrix}^T$, $\alpha_l(k) = A_l e^{j\phi'_l(k)}$, and

$$\phi'_l(k) = \phi_l - kl\omega_0 f_s dc^{-1}\sin\theta = b_l + a_l k, \tag{13}$$

with $b_l = \phi_l$ and $a_l = -l\omega_0 f_s dc^{-1}\sin\theta$. Equipped with complex amplitude estimates of the lth harmonic from all sensors in the ULA, we can estimate b_l and a_l using weighted LS (WLS) as

$$\begin{bmatrix}\hat{b}_l & \hat{a}_l\end{bmatrix}^T = \left(\mathbf{K}^T\mathbf{W}\mathbf{K}\right)^{-1}\mathbf{K}^T\mathbf{W}\boldsymbol{\phi'}_l, \tag{14}$$

where

$$\mathbf{K} = \begin{bmatrix}1 & 1 & \cdots & 1\\ 0 & 1 & \cdots & K-1\end{bmatrix}^T,$$

$\boldsymbol{\phi'}_l = \begin{bmatrix}\phi'_l(0) & \cdots & \phi'_l(K-1)\end{bmatrix}^T$, and $\mathbf{W}$ is a weighting matrix accounting for the phase estimates having different variances across the different sensors. For $\hat{b}_l$ and $\hat{a}_k$ to be statistically efficient estimates, the weighting should be $\mathbf{W}_{\boldsymbol{\phi}} = \mathbf{C}_{\boldsymbol{\phi}}^{-1}$, where $\mathbf{C}_{\boldsymbol{\phi}}$ is the covariance matrix of $\boldsymbol{\phi'}_l$. Then, an estimate of the DOA of the lth harmonic can be obtained as

$$\hat{\theta}_l = \sin^{-1}\left(\frac{-\hat{a}_l c}{l\omega_0 f_s d}\right). \tag{15}$$

The DOA of the periodic source is then estimated by combining the DOA estimates of the individual harmonics. This can be achieved by again using a WLS approach as

$$\hat{\theta} = \left(\mathbf{1}^T\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{1}\right)^{-1}\mathbf{1}^T\mathbf{C}_{\boldsymbol{\theta}}^{-1}\hat{\boldsymbol{\theta}}, \tag{16}$$

where $\mathbf{C}_{\boldsymbol{\theta}}$ is the covariance matrix of $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}} = \begin{bmatrix}\hat{\theta}_1 & \cdots & \hat{\theta}_L\end{bmatrix}^T$. In the remainder of the paper, we denote this estimator as the weighted least squares (WLS) method.

4.1. Finding the Weighting Matrices

Below, we describe how to find the weighting matrices. If the estimates to be weighted are stastically efficient, the optimal weighting matrices are simply given by the corresponding Fisher information matrices (FIMs). These matrices, however, often requires knowledge about all parameters in our signal model, which would limit the practicability of the weighting-based estimator proposed in Sec. 4. Therefore, we consider asymptotic expressions in this section to lighten this requirement.

Under the assumed noise conditions, the FIM $\mathbf{I}(\boldsymbol{\nu})$ for a vector of unknown parameters $\boldsymbol{\nu}$ is given by

$$\mathbf{I}(\boldsymbol{\nu}) = 2\mathrm{Re}\left\{\sum_{n=0}^{N-1}\frac{\partial\mathbf{s}^H(n)}{\partial\boldsymbol{\nu}}\mathbf{Q}^{-1}\frac{\partial\mathbf{s}(n)}{\partial\boldsymbol{\nu}^T}\right\}, \tag{17}$$

where

$$[\mathbf{s}(n)]_k = \sum_{l=1}^{L} A_{l,k}e^{jl\omega_0 n + j\phi'_l(k)}, \tag{18}$$

and $A_{l,k} = \beta_k A_l$. That is, it can be shown that the FIM $\mathbf{I}(\boldsymbol{\phi'}_l)$ for the phases $\boldsymbol{\phi'}_l$ is proportional to

$$[\mathbf{I}(\boldsymbol{\phi'}_l)]_{pq} \propto [\mathbf{W}_{\boldsymbol{\phi}}]_{pq}\begin{cases}\frac{A_{l,k}^2}{\sigma_k^2}, & \text{for } p=q=k,\\ (\approx)0, & \text{for } p\neq q\end{cases} \tag{19}$$

where $\mathbf{W}_{\boldsymbol{\phi}}$ can be chosen as the optimal weighting matrix in (14). As appearing from the above expression, $\mathbf{W}_{\boldsymbol{\phi}}$ only depends on the attenuated real amplitudes $A_{l,k}$ and the noise variances σ_k^2. These parameters are most likely unknown in practice, but they can easily be estimated as shown in [16]. Moreover, it can be shown that the FIM $\mathbf{I}(\boldsymbol{\theta})$ for the DOAs of the individual harmonics in $\boldsymbol{\theta}$ is proportional to

$$[\mathbf{I}(\boldsymbol{\theta})]_{pq} \propto [\mathbf{W}_{\boldsymbol{\theta}}]_{pq} = \begin{cases} l^2\cos^2\hat{\theta}_l\sum_{k=0}^{K-1}\frac{k^2A_{l,k}^2}{\sigma_k^2}, & \text{for } p=q=l,\\ (\approx)0, & \text{for } p\neq q.\end{cases} \tag{20}$$

The matrix $\mathbf{W}_{\boldsymbol{\theta}}$ can then be chosen as the optimal weighting for the estimator in (16). The weighting matrix $\mathbf{W}_{\boldsymbol{\theta}}$ depends on the attenuated real amplitudes $A_{l,k}$, the noise variances σ_k^2, and the DOAs of the individual harmonics θ_l. The amplitudes and noise variances are estimated easily in practice as previously described, while θ_l can be estimated using (15). If desired, we can further simplify the expression in (20) by assuming that the individual DOAs of the harmonics are nearly the same in which case the $\cos^2\theta_l$ term can be left out.

5. EXPERIMENTAL RESULTS

Next, we present the experimental evaluation of the proposed methods. First, we conducted several Monte-Carlo simulations where we evaluated the methods proposed in Section 3 (JML) and 4 (WLS), respectively, as well as other methods for comparison. The other methods in this comparison are the maximum likelihood pitch estimator (ML) in [16], the steered response power method (SRP) [28], the position-pitch plane based method (POPI) in [21], and the nonlinear least squares method (NLS) in [25]. Note that we used FFT lengths of 256 and 1,024 in our SRP and POPI implementations, respectively, and in the SRP method, we integrated over all frequency indices. In each of these simulations, the sampling frequency was 8 kHz, the wave propagation speed was $c = 343$ m/s, and a uniform linear array was used with sensor spacing $d = c/f_s$. Arrays containing up to 10 sensors were considered, where the SNRs and attenuation factors on the different sensors were $[40, 20, 10, 25, 15, 20, 30, 35, 25, 40]$ dB and $\boldsymbol{\beta} = [1, 0.99, 0.98, 0.97, 0.96, 0.95, 0.94, 0.93, 0.92, 0.91]^T$, respectively. Moreover, a periodic signal in white Gaussian noise was considered in each simulation, where the periodic signal consisted of $L = 4$ harmonics with amplitudes $|\boldsymbol{\alpha}| = [1, 0.8, 0.6, 0.2]^T$ and random, uniformly distributed phases, and the signal had a pitch of $f_0 = 243$ Hz and a DOA of $\theta = -15^\circ$ onto the array. Finally, 500 Monte-Carlo simulations were conducted for each parameter setting. In the first series of Monte-Carlo simulations, the mean squared error (MSE) of the pitch and DOA estimates were measured as a function of the average SNR across the sensors in the array. In this series, the first $K = 4$ sensors of the ULA was used, the number of temporal samples was $N = 60$, and the average SNRs were obtained by scaling the aforementioned sensor SNRs. Furthermore, we conducted a series of simulations, where the number of sensors K was varied and the number of snapshots was $N = 60$, and, finally, we measured the performance for different number of temporal samples, when the number of sensors was $K = 4$. The outcome of these evaluations are depicted in Fig. 1. From these results, we first observe that there is only a subtle difference between using the exact CRBs calculated from the true parameter values (C-WLS) and the asymptotic CRBs based on estimated parameter values (WLS) as the weights for the method in Sec. 4. We also observe that the proposed methods outperforms all other methods in the considered

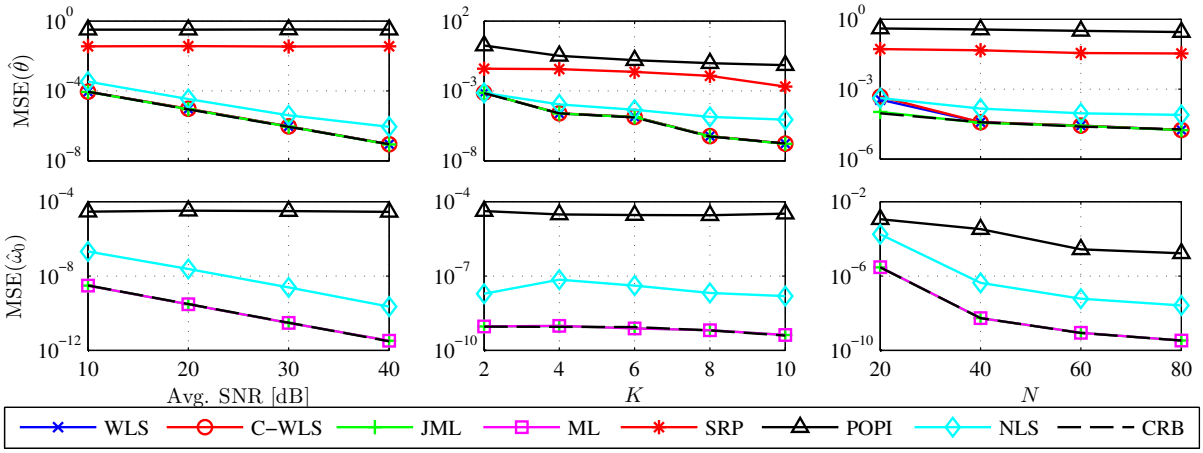

Fig. 1. Mean squared errors of DOA and pitch estimates obtained using various methods in different scenarios.

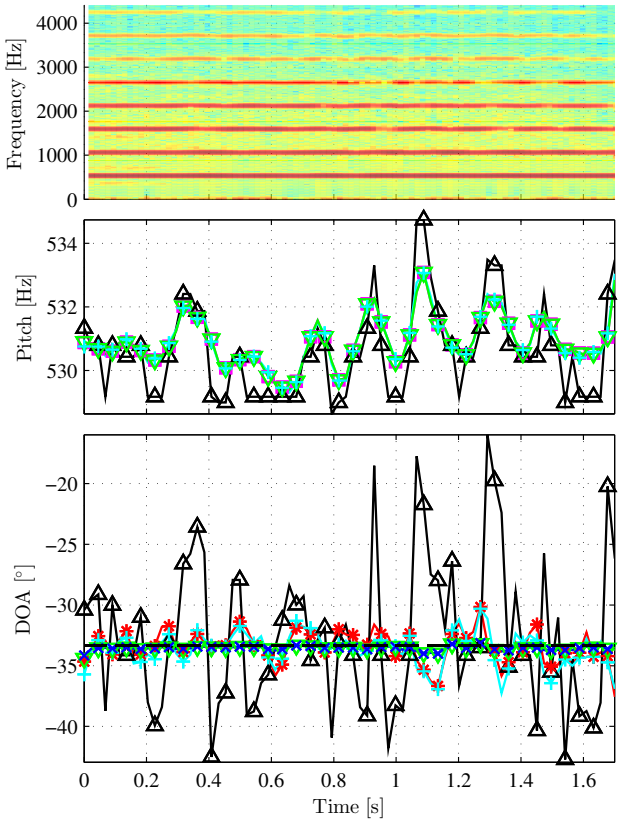

Fig. 2. Plots of (upper) the spectrogram of the utilized trumpet signal, (middle) the pitch estimates, and (bottom) the DOA estimates. The labels for the two lower plots are found in Fig. 1.

scenarios, and that they attain the Cramér-Rao bounds as expected. For very low numbers of samples, the proposed JML method seems to outperform the proposed WLS method for DOA estimation.

We also evaluated the aforementioned methods for estimation of the pitch and DOA of a multichannel, real-life signal. The multichannel signal was obtained by synthesizing a single-channel trumpet signal spatially by using the image method [29]. A spectrogram of the trumpet signal can be found in Fig. 2. For generating the multichannel signal in this way, we used an online MATLAB implementation [30] of the image method and the setup was as follows: the length of the acoustical room impulse responses (RIRs) was 4,096, cardioid microphones were used, the reflection order was 0, the room dimension was 3 and the dimensions were $5\times4\times3$ m, the microphone orientation was 0, and the highpass filter was disabled. Furthermore, four microphones with coordinates (0.2, $1-1.5d$, 2) m, (0.2, $1-0.5d$, 2) m, (0.2, $1+0.5d$, 2) m, and (0.2, $1+1.5d$, 2) m were used with $d = f_s/c$, and the source was placed at (4, 3.5, 2) m; this corresponds to a DOA of $\approx$33.3°. The SNRs for the four microphones were [40, 10, 40, 0] dB, while the speed of sound and the sampling frequency were the same as in the previous experiment. As we do not consider model order estimation, the model order was fixed to $L = 4$. Using this setup, we applied the aforementioned methods on blocks of $N = 200$ samples of the generated, multichannel signal, and the resulting estimates over time are depicted in Fig. 2. From these results, we first observe that the obtained pitch estimates are consistent with the spectrogram of the trumpet signal. Moreover, we observe that the JML, ML and NLS methods yield pitch estimates with similar variance, whereas the variance is somewhat larger for the estimates obtained using the POPI method. For DOA estimation, the proposed WLS and JML methods provide estimates close to the true DOA and they seem to outperform all the other methods in the comparison (SRP, NLS and POPI) with the POPI method having the worst performance.

6. DISCUSSION

The work in this paper is considering the topic of estimation of the pitch and DOA of multichannel, periodic sources. Only a few methods for estimating these parameters jointly have been proposed including the maximum-likelihood based method in [17], the subspace methods in [18–20], the correlation-based methods in [21, 22], the filtering methods in [23, 24], and the nonlinear least squares based methods in [25]. Some of these methods estimate the time delay instead of the DOA, however, these parameters are closely related. To the best of our knowledge, none of these existing methods yield maximum likelihood estimates when the noise is white Gaussian and the SNRs are different across the channels. However, this can be achieved with the methods proposed herein, which is also clear from the reported results. Furthermore, the proposed methods show superior performance compared to the other state-of-the-art methods when applied on a real-life signal.

7. REFERENCES

[1] H. Wang and P. Chu, "Voice source localization for automatic camera pointing system in videoconferencing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 1997, vol. 1, pp. 187–190.

[2] D. Chazan, Y. Stettiner, and D. Malah, "Optimal multi-pitch estimation using the EM algorithm for co-channel speech separation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 1993, vol. 2, pp. 728–731.

[3] M. G. Christensen and A. Jakobsson, "Optimal filter designs for separating and enhancing periodic signals," *IEEE Trans. Signal Process.*, vol. 58, no. 12, pp. 5969–5983, Dec. 2010.

[4] J. R. Jensen, M. G. Christensen, and S. H. Jensen, "An optimal spatio-temporal filter for extraction and enhancement of multi-channel periodic signals," in *Rec. Asilomar Conf. Signals, Systems, and Computers*, Nov. 2010, pp. 1846–1850.

[5] J. R. Jensen, J. Benesty, M. G. Christensen, and S. H. Jensen, "Enhancement of single-channel periodic signals in the time-domain," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 20, no. 7, pp. 1948–1963, Sep. 2012.

[6] H. Purnhagen and N. Meine, "HILN - the MPEG-4 parametric audio coding tools," in *Proc. IEEE Int. Symp. Circuits and Systems*, May 2000, vol. 3, pp. 201–204.

[7] J. Lindblom, "A sinusoidal voice over packet coder tailored for the frame-erasure channel," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 5, pp. 787–798, Sep. 2005.

[8] C. H Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 24, no. 4, pp. 320–327, Aug. 1976.

[9] H. Wang and M. Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 4, pp. 823–831, Aug. 1985.

[10] G. C. Carter, "Coherence and time delay estimation," *Proc. IEEE*, vol. 75, no. 2, pp. 236–255, Feb. 1987.

[11] M. S. Brandstein and H. F. Silverman, "A robust method for speech signal time-delay estimation in reverberant rooms," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 1997, vol. 1, pp. 375–378.

[12] M. Jian, A. C. Kot, and M. H. Er, "DOA estimation of speech source with microphone arrays," in *Proc. IEEE Int. Symp. Circuits and Systems*, May 1998, vol. 5, pp. 293–296.

[13] M. G. Christensen and A. Jakobsson, "Multi-pitch estimation," *Synthesis Lectures on Speech and Audio Processing*, vol. 5, no. 1, pp. 1–160, 2009.

[14] L. Armani and M. Omologo, "Weighted autocorrelation-based f0 estimation for distant-talking interaction with a distributed microphone network," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2004, vol. 1, pp. 113–116.

[15] F. Flego and M. Omologo, "Robust f0 estimation based on a multi-microphone periodicity function for distant-talking speech," in *Proc. European Signal Processing Conf.*, Sep. 2006, pp. 1–4.

[16] M. G. Christensen, "Multi-channel maximum likelihood pitch estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2012, pp. 409–412.

[17] X. Qian and R. Kumaresan, "Joint estimation of time delay and pitch of voiced speech signals," *Rec. Asilomar Conf. Signals, Systems, and Computers*, vol. 1, pp. 735–739, Oct. 1995.

[18] G. Liao, H. C. So, and P. C. Ching, "Joint time delay and frequency estimation of multiple sinusoids," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2001, vol. 5, pp. 3121–3124.

[19] L. Y. Ngan, Y. Wu, H. C. So, P. C. Ching, and S. W. Lee, "Joint time delay and pitch estimation for speaker localization," in *Proc. IEEE Int. Symp. Circuits and Systems*, May 2003, vol. 3, pp. 722–725.

[20] J. X. Zhang, M. G. Christensen, S. H. Jensen, and M. Moonen, "Joint DOA and multi-pitch estimation based on subspace techniques," *EURASIP J. on Advances in Signal Processing*, vol. 2012, no. 1, pp. 1–11, Jan. 2012.

[21] M. Wohlmayr and M. Képesi, "Joint position-pitch extraction from multichannel audio," in *Proc. Interspeech*, Aug. 2007, pp. 1629–1632.

[22] M. Képesi, L. Ottowitz, and T. Habib, "Joint position-pitch estimation for multiple speaker scenarios," in *Proc. Hands-Free Speech Commun. Microphone Arrays*, May 2008, pp. 85–88.

[23] J. Dmochowski, J. Benesty, and S. Affes, "Linearly constrained minimum variance source localization and spectral estimation," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 16, no. 8, pp. 1490–1502, Nov. 2008.

[24] J. R. Jensen, M. G. Christensen, and S. H. Jensen, "Joint DOA and fundamental frequency estimation methods based on 2-d filtering," in *Proc. European Signal Processing Conf.*, Aug. 2010, pp. 2091–2095.

[25] J. R. Jensen, M. G. Christensen, and S. H. Jensen, "Nonlinear least squares methods for joint DOA and pitch estimation," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 21, no. 5, pp. 923–933, May 2013.

[26] G. L. Bretthorst, "An introduction to parameter estimation using Bayesian probability theory," in *Maximum Entropy and Bayesian Methods*, P. Fougere, Ed., pp. 53–79. Kluwer Academic Publishers, 1990.

[27] E.-W. Bai and Y. Liu, "Least squares solutions of bilinear equations," *Systems & Control Lett.*, vol. 55, no. 6, pp. 466–472, 2006.

[28] J. H. DiBiase, H. F. Silverman, and M. S. Brandstein, "Robust localization in reverberant rooms," in *Microphone Arrays - Signal Processing Techniques and Applications*, M. S. Brandstein and D. B. Ward, Eds., chapter 8, pp. 157–180. Springer-Verlag, 2001.

[29] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.

[30] E. A. P. Habets, "Room impulse response generator," Tech. Rep., Technische Universiteit Eindhoven, 2010, Ver. 2.0.20100920.